AD_____

Award Number: DAMD17-99-1-9119

TITLE: A New Model for the Estimation of Breast Cancer Risk

PRINCIPAL INVESTIGATOR: Maryellen Giger, Ph.D.

CONTRACTING ORGANIZATION: The University of Chicago
Chicago, Illinois 60637

REPORT DATE: July 2000

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	July 2000	Annual (14 Jun 99 - 13 Jun 00)

4. TITLE AND SUBTITLE
A New Model for the Estimation of Breast Cancer Risk

5. FUNDING NUMBERS
DAMD17-99-1-9119

6. AUTHOR(S)
Maryellen Giger, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
The University of Chicago
Chicago, Illinois 60637

E-MAIL:
m-giger@uchicago.edu

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT ***(Maximum 200 Words)***

Cancer risk is the probability that cancer will occur in a given population. Research on cancer risk seeks to identify populations with a high probability of developing cancer. The goal of this research is to merge a computerized analysis of mammograms, which characterizes the breast pattern, with information of a woman's personal and family histories into a novel model for use in estimating risk of breast cancer. The specific aims include 1. Creating a database of mammograms, along with tabulated clinical information of women at low risk and high risk for breast cancer; 2. Developing a new model using computer methods for merging mammographic information with clinical information; and 3. Evaluating the efficacies of the new model compared to currently used methods of risk assessment. The main hypothesis to be tested is that given a group of women, the new computerized risk model that merges computerized analyses of mammograms with clinical information should yield a novel way for identifying those women at risk for breast cancer.

The new model will include computer-extracted features from digitized mammograms and clinical information from each woman. The computer-extracted features will be extracted within regions of digitized mammograms. In general, the breast can be described by the amount of dense regions (a percent dense) and by the heterogeneity/homogeneity of the dense portion pattern (texture). In addition, clinical information such as age and reproductive history contribute to the determination of risk. Therefore, methods of combining clinical data and multiple mammographic markers into a single model of risk will be developed for the model.

14. SUBJECT TERMS
Breast Cancer

15. NUMBER OF PAGES
9

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

FOREWORD

Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the U.S. Army.

___ Where copyrighted material is quoted, permission has been obtained to use such material.

___ Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

___ Citations of commercial organizations and trade names in this report do not constitute an official Department of Army endorsement or approval of the products or services of these organizations.

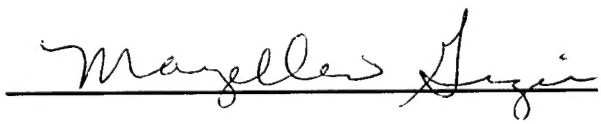
__ In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and use of Laboratory Animals of the Institute of Laboratory Resources, national Research Council (NIH Publication No. 86-23, Revised 1985).

X For the protection of human subjects, the investigator(s) adhered to policies of applicable Federal Law 45 CFR 46.

N/A In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institutes of Health.

N/A In the conduct of research utilizing recombinant DNA, the investigator(s) adhered to the NIH Guidelines for Research Involving Recombinant DNA Molecules.

N/A In the conduct of research involving hazardous organisms, the investigator(s) adhered to the CDC-NIH Guide for Biosafety in Microbiological and Biomedical Laboratories.



Proposal Title: A New Model for the Estimation of Breast Cancer Risk

P.I.: Maryellen L. Giger, Ph.D.

TABLE OF CONTENTS

INTRODUCTION:

Cancer risk is the probability that cancer will occur in a given population. Research on cancer risk seeks to identify populations with a high probability of developing cancer. The goal of this research is to merge a computerized analysis of mammograms, which characterizes the breast pattern, with information of a woman's personal and family histories into a novel model for use in estimating risk of breast cancer.

The specific aims include 1. Creating a database of mammograms, along with tabulated clinical information of women at low risk and high risk for breast cancer; 2. Developing a new model using computer methods for merging mammographic information with clinical information; and 3. Evaluating the efficacies of the new model compared to currently used methods of risk assessment. The main hypothesis to be tested is that given a group of women, the new computerized risk model that merges computerized analyses of mammograms with clinical information should yield a novel way for identifying those women at risk for breast cancer. It should be noted that current clinical methods of assessing risk using the Gail or Claus models (clinical data only) are limited as illustrated by our preliminary studies, which show only moderate correlation between these two current models for cumulative risk and 10-year risk.

The new model will include computer-extracted features from digitized mammograms and clinical information from each woman. The computer-extracted features will be extracted within regions of digitized mammograms. In general, the breast can be described by the amount of dense regions (a percent dense) and by the heterogeneity/homogeneity of the dense portion pattern (texture). In addition, clinical information such as age and reproductive history contribute to the determination of risk. Therefore, methods of combining clinical data and multiple mammographic markers into a single model of risk will be developed for the model.

Potential uses of this innovative model include 1) serving as a means to assess the cancer risk of women undergoing routine screening mammography and thus, identifying those women that may require closer scrutiny and 2) serving as a means to monitor the cancer risk of women undergoing chemoprevention treatments. The research is novel in that currently there does not exist a reliable means to assess the cancer risk of individual women using both mammographic and clinical information. In addition, if a woman knew that she was at an increased risk of breast cancer, it is likely that she would better comply with screening mammography programs. In the future, a successful model could also be used to assess the effect of chemoprevention on a women's parenchymal pattern and thereby, overall risk.

BODY:

Task 1. Establishment of database (mos. 1-30)

The high-risk database is being collected within the University of Chicago Cancer Risk Clinic and consists of mammograms, pedigree information, epidemiological data and related biological specimens from patients with a family history of breast cancer. All mammograms done since 1990 are being collected for all participants irrespective of their cancer status. Breast Cancer risk assessment is performed using both Gail and Claus models and genetic testing whenever possible. A low-risk database is also being collected from our breast cancer screening program and includes mammograms and clinical information on women undergoing routine screening mammograms. The low risk database is being developed to include women who are age-matched to reflect the age of women in our high risk database. We have collected cases from over 100 patients and Gail and Claus calculations have been performed. We now have approximately 35 patients with positive BRCA1/BRCA2 gene mutation testing.

The mammograms are converted to digital format by using a laser film scanner (2048 by 2048 matrix with 12-bit quantization). Such high spatial resolution is necessary in order to adequately retain the high-frequency texture patterns.

Task 2. Development of risk model including mammographic markers and clinical information (mos. 3-30)

Computerized analysis of the parenchymal pattern is based on various texture analysis methods we have developed in our laboratory including Fourier spectra analysis, histogram analysis, and artificial neural networks. Fourteen features are currently extracted within the regions of each digitized mammogram. These features are grouped into (i) features based on the absolute values of the gray levels, (ii) features based on gray-level histogram analysis, (iii) features based on the Fourier transform, and (iv) features based on the spatial relationship among gray levels.

The purpose of one of our studies, was to identify computer-extracted, mammographic parenchymal patterns that are associated with breast cancer risk. We extracted fourteen features from the central breast region on digitized mammograms to characterize the mammographic parenchymal patterns of women at different risk levels. In the study, the features were used to characterize mammographic patterns seen in low-risk women and in women who have breast cancer. Stepwise linear logistic regression was employed to identify useful features to differentiate between the mammographic

patterns of low-risk women and women with breast cancer. The relationship between these mammographic patterns and the risk of developing breast cancer was identified based on the odds ratios associated with these individual features. This analysis is considered to be the third approach with which we relate our computer-extracted mammographic features to breast cancer risk. Previously, we employed two different approaches to relate these mammographic features to breast cancer risk. In one approach, the features were used to distinguish mammographic patterns seen in low-risk women from those who inherited a mutated form of the *BRCA1/BRCA2* gene. In another approach, the features were related to risk as determined from existing clinical models (*Gail* and *Claus* models). Stepwise linear discriminant analysis was employed to identify features that were useful in differentiating between "low-risk" women and *BRCA1/BRCA2*-mutation carriers. Stepwise linear regression analysis was employed to identify useful features in predicting the risk as estimated from the *Gail* and *Claus* models. The computer-extracted mammographic features identified from this approach were similar to those identified from the two previous approaches. The results from this study show that women who have dense breasts and whose mammographic patterns are coarse and low in contrast have an increased risk of developing breast cancer. The consensus of the findings from the three different approaches substantiated the existing results. (Presented CARS 2000)

We also analyzed the contributions of age and computer-extracted mammographic features in the prediction of breast cancer risk. We assessed the contribution of the computer-extracted features to risk prediction in terms of percent increase in the prediction power (r2) when age (the single most important risk factor for breast cancer) was used alone and when the mammographic features were included. The inclusion of the mammographic features increased the prediction power (r2) from 0.08 and 0.16 (age alone) to 0.17 and 0.32, yielding an increase of 113% an d 100% in r2 for predicting the risk as estimated from the Gail and Claus models. The substantial increase in r2 indicates the important contribution of these mammographic features in risk prediction and the need to incorporate in predicting breast cancer risk. (Presented IWDM 2000)

Task 3. Evaluation methods

This task is planned for months 20-36. However, ourplans include the following.

We are developing a model for assessing breast structure and cancer risk. Thus, correlation analysis will be used in evaluating the performance of the measures. Linear correlation analysis will be performed to determine the correlation among the output of the new model and the Gail risk model (or Claus model).

Another task in which the combined measures will be evaluated will be in their ability to predict the onset of breast cancer (over time). Based on the cases collected during the first 2.5 years of the study, a nested case-control study design will be implemented. As our criteria are that the mammograms should have been obtained after 1989, there is potential for collecting images from eight years ago (so can assume 5 to 8 year follow-up). In a nested case-control database (36), the cases will correspond to women who will have developed cancer and the control will correspond to women who will have stayed cancer free during the period. We will calculate the clinical markers (e.g., Gail) and the mammographic features of the initial examination prior to the 5 to 8 year follow-up. Multivariate analysis will be used to examine the relationship between the new model and risk of breast cancer while controlling for other risk factors such as age at menarche and parity. A proportional-hazards regression model will be used to calculate the relative risk for each radiographic marker.

KEY RESEARCH ACCOMPLISHMENTS:

- Increase in our database of high and low risk cases, especially those with positive BRCA1/BRCA2 testing.
- Further verification of our texture features for characterizing the breast parenchyma using a third approach
- Preliminary study looking at the contribution of age and mammographic features to breast cancer risk prediction.

REPORTABLE OUTCOMES:

1. International Workshop on Digital Mammography 2000 (Toronto, Canada)

   Analysis of the relative contributions of mammographic features and age to breast cancer risk prediction. Zhimin Huo, Maryellen L. Giger and Olufunmilayo I. Olopade

2. CARS 2000 (San Fransico, CA)

   Computerized analysis of mammographic patterns of women with and without breast cancer. Zhimin Huo, Maryellen L. Giger and Olufunmilayo I. Olopade

CONCLUSIONS:

To date, we have shown that computer-extracted features of mammographic parenchymal patterns can be used in the prediction of breast cancer risk. This has been demonstrated (on the developing database) using three approaches: (1) correlation with clinical models of Gail and Claus, (2) separation between women at low risk and those with a positive gene testing result, and (3) separation between women at low risk and those that have breast cancer. In addition, we have shown, in a preliminary study, that the inclusion of the mammographic features with age increase the predictive power over the use of age alone in the prediction of breast cancer risk.